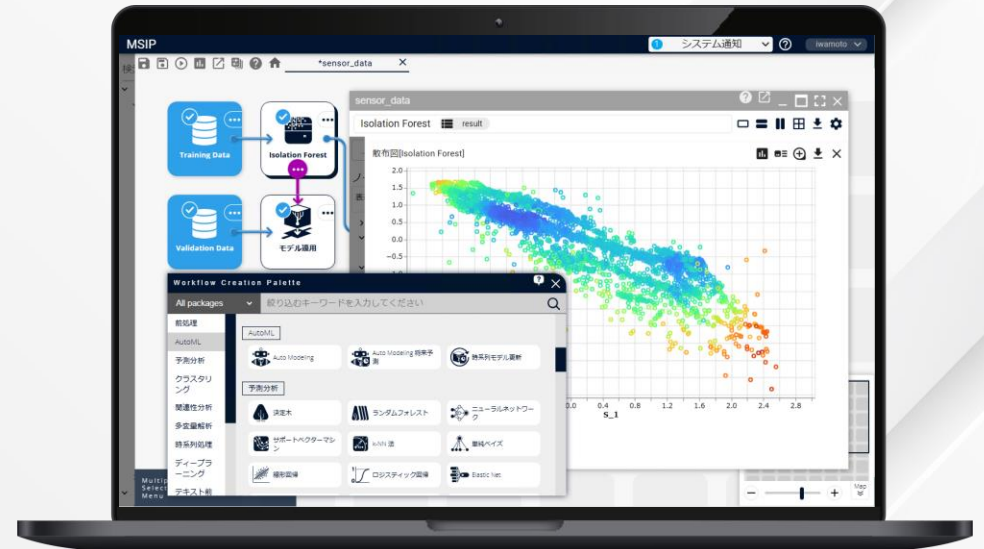


テクニカルサンプルプロジェクト 時系列クラスタリング分析



株式会社 NTTデータ数理システム

このプロジェクトについて

こんな方におすすめします

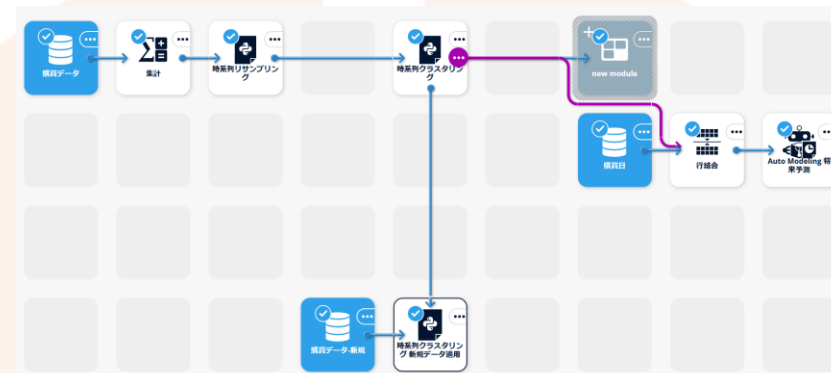
- ✓ たくさんの時系列の間に潜む共通の 패턴を見出し、データの理解に役立てたい方
- ✓ 時系列データを機械的に分類・整理するための時系列クラスタリング手法に興味がある方

何をするプロジェクト？

ビジネスでは、小売店の購買データや製造設備のセンサーデータなど、様々な時系列データが日々蓄積されています。しかし、たくさんの時系列データを単に眺めるだけでは、有用な知見を得るのは困難です。

本プロジェクトでは、時系列クラスタリングを用いることで、複数の時系列を系列間にある共通パターンからグループ化する方法をご紹介します。さらに、時系列データのグループ化と「AutoModeling 将来予測」アイコンによる分析を組み合わせることで、抽出したパターンの将来動向を予測する例も紹介します。

例えば、小売店の「顧客購買データ」を対象にした本プロジェクトの分析では、顧客の購買行動に基づくグループ化や、行動パターンごとの商品購入予測を通じて、有効なマーケティング施策を検討することが可能です。

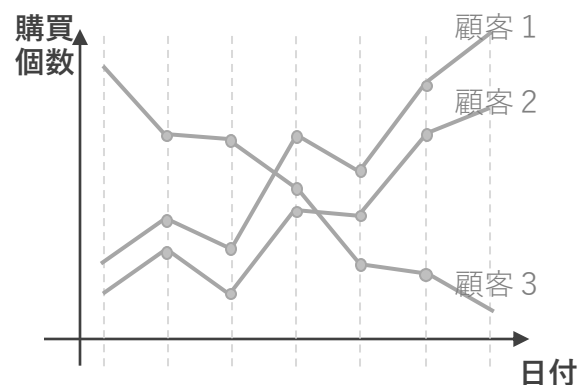


プロジェクトのポイント

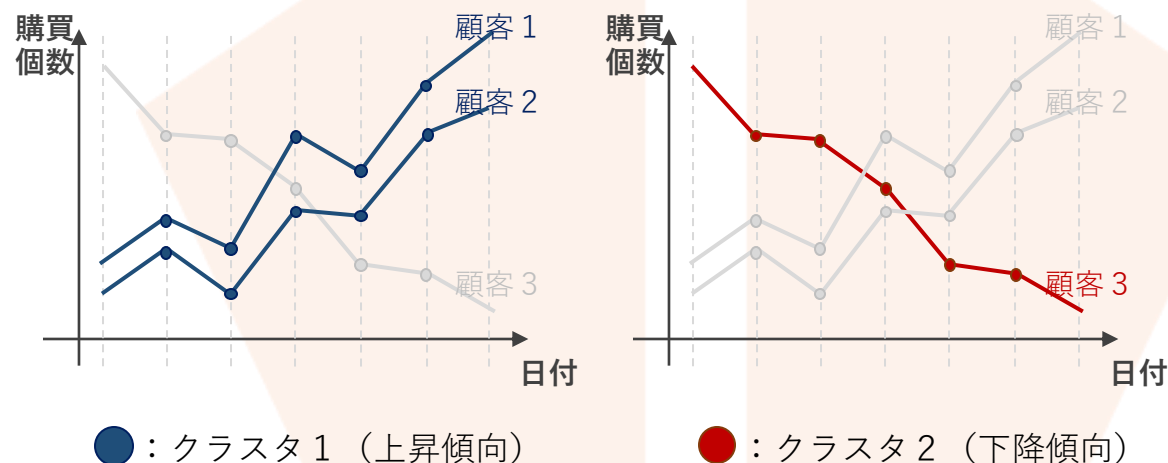
本プロジェクトでは、小売店の「顧客購買データ」に時系列クラスタリングを適用し、購買パターンに基づく顧客のグループ分け（クラスタリング）を行います。

時系列クラスタリングで顧客をグループ化するイメージ

クラスタリング前



クラスタリング後



プロジェクト 分析フロー概要

本プロジェクトは、大きく4つのフェイズから構成されます。

本プロジェクトにおける分析フロー

データの確認

- ✓ プロジェクトで用いる「購買データ」と「顧客マスタ」の確認をします。この部分では、主に購買データに焦点を当ててデータの傾向を把握していきます。

前処理

- ✓ 時系列リサンプリングを用いて、購買データの時間間隔が等間隔になるよう変換します。あわせて、購買データが「顧客ごとの時系列」の形式になるよう整形します。

時系列クラスタリング

- ✓ 時系列クラスタリングによって、購買パターンが似ている顧客をグループ化します。また、各顧客グループの購買パターンをプロットによって確認します。

追加的な分析

- ✓ クラスタリングによって得られた各グループが、どのような属性の顧客で構成されているかを確認します。
- ✓ また、クラスタ中心線に対して時系列予測を行い、各グループの購買傾向が今後どのように推移するかを予測します。

プロジェクト 解説

プロジェクト 解説

1. 対象データ

■ 購買データ

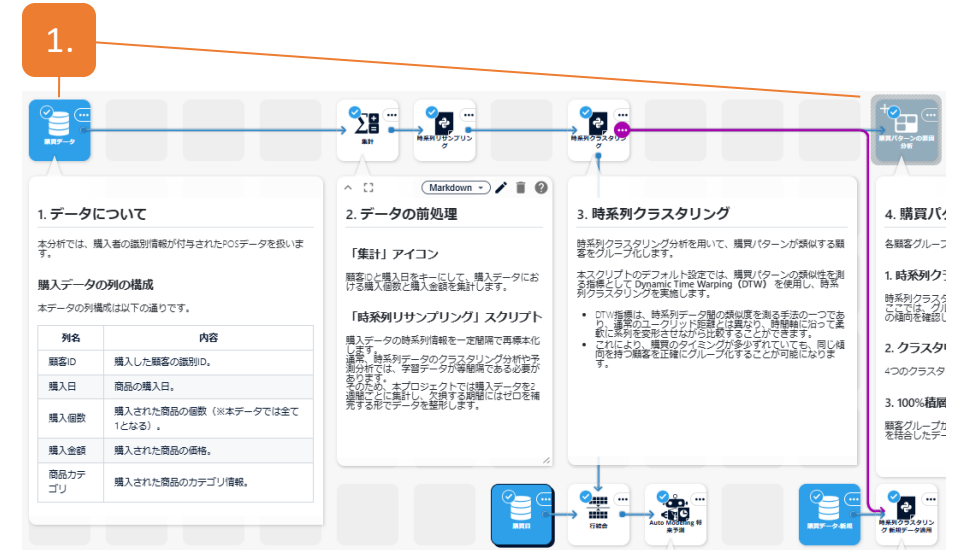
- 小売店における顧客の購買情報を時系列順に記録したデータです。

顧客ID	購買日	購買個数	購買金額	商品カテゴリ
3	2020/01/01	1	1,900	clothing
13	2020/01/01	1	100	book
16	2020/01/01	1	1,600	book
...	...	...	...	...

■ 顧客マスタ

- 購買データに含まれる、各顧客の属性情報データです。

顧客ID	性別	年代
0	F	60
1	M	50
2	M	40
...	...	...



プロジェクト 解説

2. データの前処理

■ 集計

- 顧客IDと購買日をキーにして、購買データにおける購買個数と購買金額を集計します。

■ 時系列リサンプリング

- 時系列データのクラスタリング分析および予測分析では、学習データは通常、等間隔なデータである必要があります。
- 本プロジェクトでは、系列データを再標本化することで、この形式のデータへ変換します。

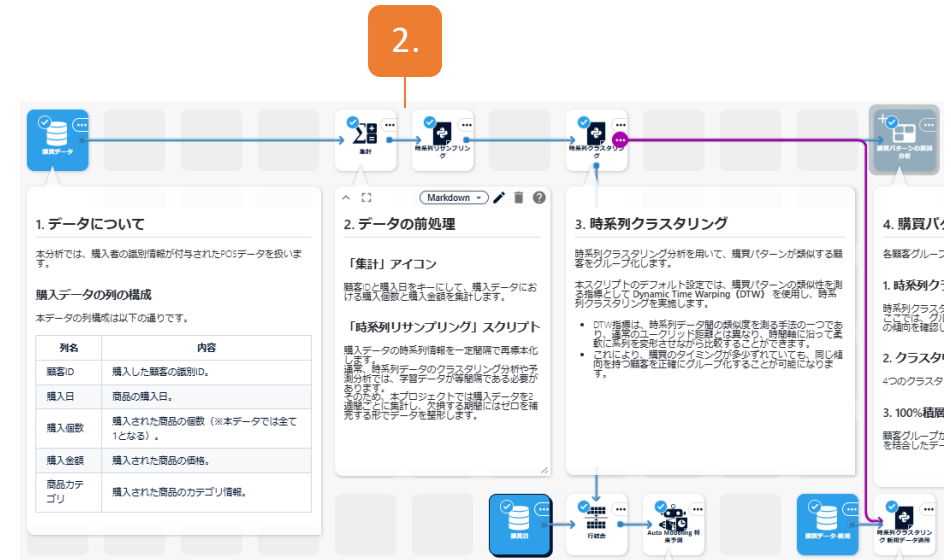
◆ 等間隔（2週間おき）に再標本するイメージ

顧客ID	購買日	購買個数(合計)	購買金額(合計)
0	2024/11/25	2	3000
0	2024/11/28	1	1000
0	2024/12/2	7	7500
0	2024/12/5	2	2500
⋮	⋮	⋮	⋮
1	2024/11/25	5	5000
1	2024/11/27	2	2200
⋮	⋮	⋮	⋮



顧客ID	購買日	購買個数(合計)	購買金額(合計)
0	2024/11/25	2	3000
0	2024/12/2	7	7500
0	2024/12/23	3	3500
⋮	⋮	⋮	⋮
1	2024/11/25	5	5000
1	2024/12/9	2	2200
1	2024/12/23	0	0
⋮	⋮	⋮	⋮

※データを等間隔にする際、その間隔内に値をもたない行にはゼロを補間します。



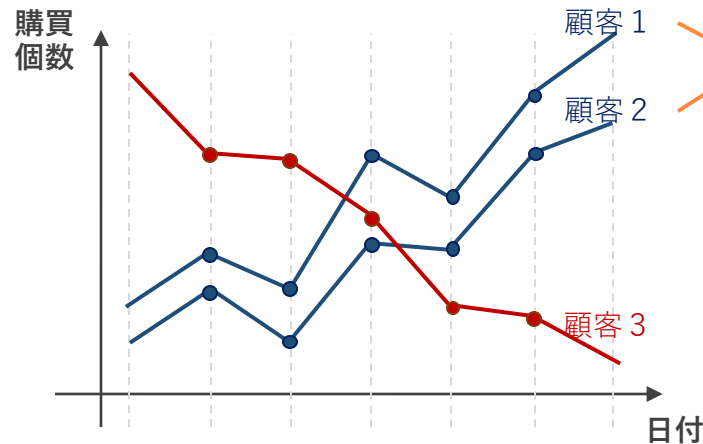
プロジェクト 解説

3. 時系列クラスタリング

■ 時系列クラスタリング

- 時系列の類似度を測る指標として、ユークリッド距離かDynamic Time Warping (DTW) 指標を選択できます。
- 本プロジェクトでは、DTW指標が選択されています。

◆ 顧客をグループ分けするイメージ



各時点での傾向が
似ている（上昇傾向）



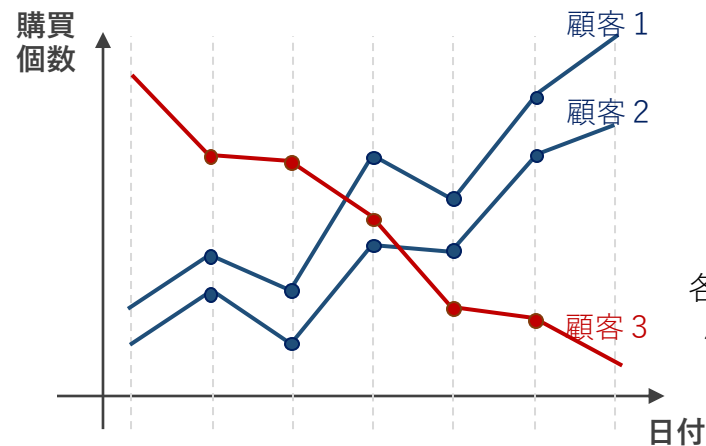
プロジェクト 解説

4. 購買パターンの要因分析

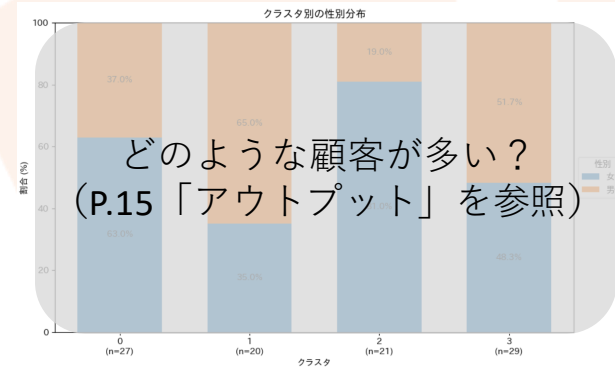
■ クラスタリング結果のグラフによる可視化

- クラスタリング結果を顧客マスタとマージした後、クラスタリングで得られた各グループがどのような属性の顧客で構成されているかを確認します。

◆ 要因分析で考えることのイメージ



各クラスターの顧客属性をグラフ化

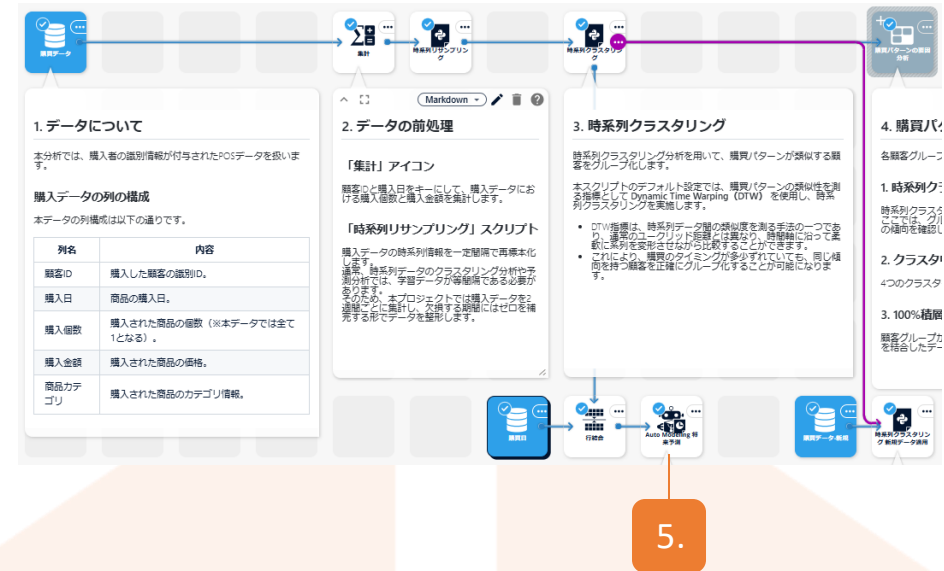


プロジェクト 解説

5. クラスタ中心線に対する時系列予測

■ AutoModeling 将来予測

- 時系列クラスタリングにおけるクラスタの中心線は、各クラスタを構成する系列の平均的な推移を表します。
- 本プロジェクトでは、顧客グループのクラスタ中心線に対して「AutoModeling 将来予測」アイコンを用いた時系列予測を行い、各顧客グループの購買行動が今後どのように推移するかを予測します。



- ✓ AutoModeling 将来予測アイコンでは、勾配ブースティング決定木（GBDT）、Prophet、ARIMA、SARIMAXの中から予測精度の観点で最適なモデルを探索し、予測を行います。
- ✓ 本プロジェクトでは、探索対象のモデルとして GBDTとProphetを指定しています。

プロジェクト 解説

6. 時系列クラスタリングへの新規データ適用

■ 時系列クラスタリング 新規データ適用

- 時系列クラスタリング（P.8 参照）のモデルに対して、新規データを適用することで、「新規データにおける各顧客がどのクラスタに属するか」を判定します。

◆ 新規データにおける顧客のクラスタ予測のイメージ

顧客ID	購買日	購買個数（合計）	購買金額（合計）
101	2021/01/05	4	26,080
101	2020/01/19	11	87,190
101	...	...	...
101	2021/12/05	7	61,550
102	2020/01/05	5	30,280
102	2020/01/19	6	29,290



cluster 2 に該当

cluster 4 に該当

新規データ内の顧客が、
どのクラスタに属するかを判定

6.

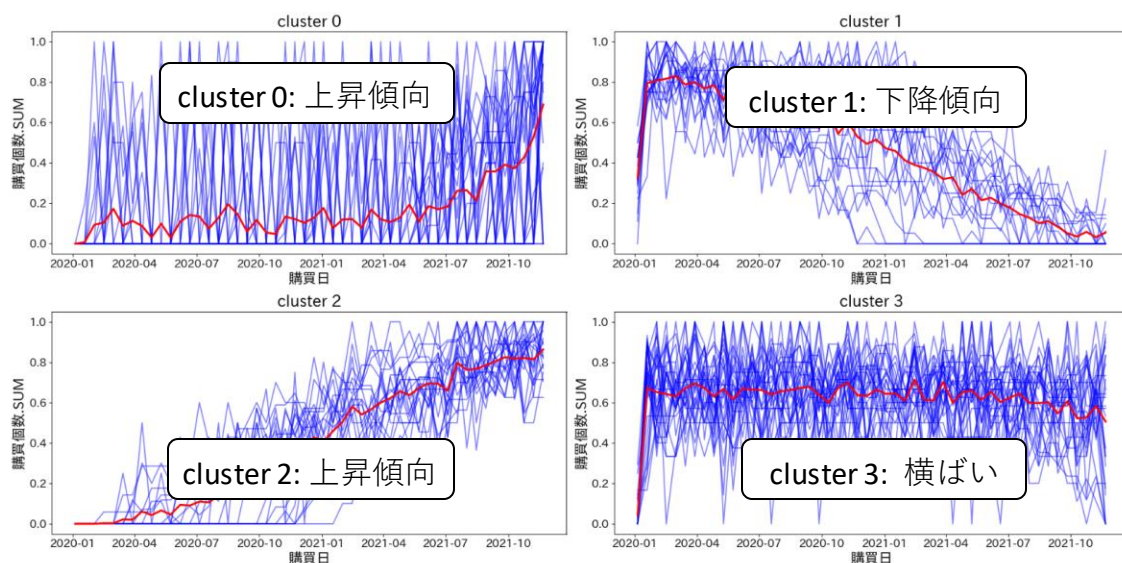
アウトプットの説明

アウトプット

購買パターンの要因分析（その1）

時系列クラスタリングでは、購買個数の推移の傾向が似ている顧客をグループ化します。ここでは、グループ数（クラスタ数）を4に指定し、各顧客グループの時系列データをプロットした図をもとにそれぞれの傾向を確認します。

・ 時系列クラスタリングによる顧客グループごとの時系列プロット（cls_preds_plot）



青線:各時点での顧客の購買個数 赤線: クラスタ中心線

プロットからの観察

- ・ 購買個数の推移から、グループごとの上昇傾向、下降傾向、横ばいといった全体的な傾向を把握できます。
- ・ グループ間で顧客数に大きな差があることも確認できます。

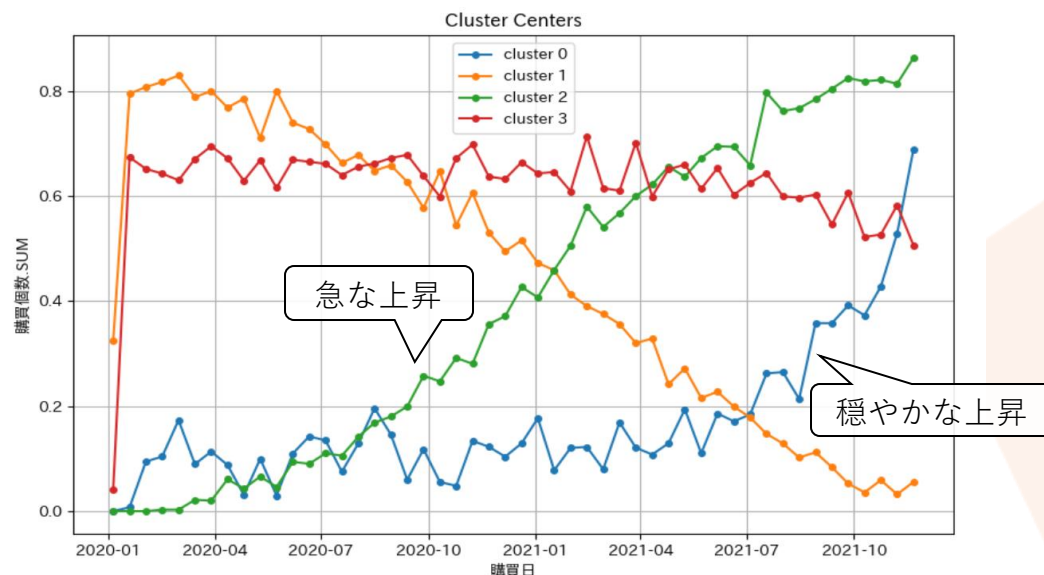
- ✓ 顧客グループごとの売上トレンドをもとに、**マーケティング施策の実施が必要かを判断**できます。
- ✓ グループ内の顧客数は、そのグループに対する**施策の重要性を判断するための参考材料**となります。

アウトプット

購買パターンの要因分析（その2）

前頁の4つのクラスタの中心線を一枚のプロットに重ね合わせた図から、グループ間の上昇・下降の程度を比較します。

- クラスタ中心線を重ね合わせたプロット（cls_centers_plot）



プロットからの観察

- 「上昇傾向」のグループ内でも、穏やかな上昇（cluster 0）と急な上昇（cluster 2）との2つのパターンがあることが確認できます。

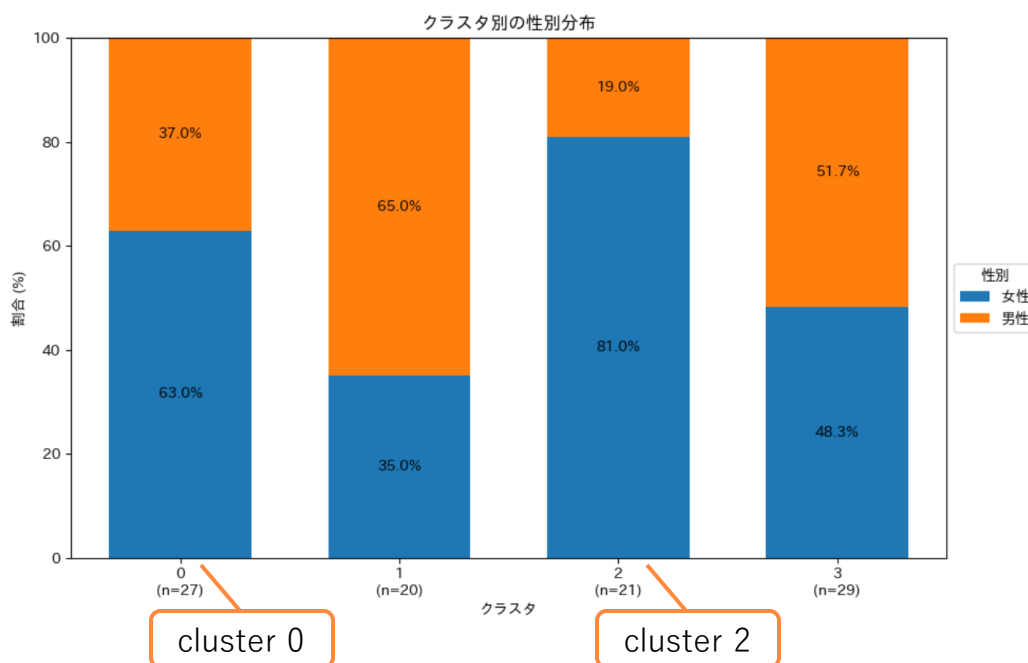
✓ 上昇傾向の違いをもとに、そのグループへのマーケティング施策の優先順位や実施のタイミングを判断できます。

アウトプット

「100%積層棒グラフ」 スクリプト（購買パターンの要因分析）

顧客グループがどのような属性の顧客で構成されているかを確認します。ここでは、クラスタリングの結果と顧客マスタを結合したデータを使用し、各クラスターの男女比率を棒グラフで可視化しています。

・ 各クラスターにおける性別の割合（plot）



プロットからの観察

- cluster 0（穏やかな上昇）とcluster 2（急な上昇）では、女性の割合が高いことを確認できます。

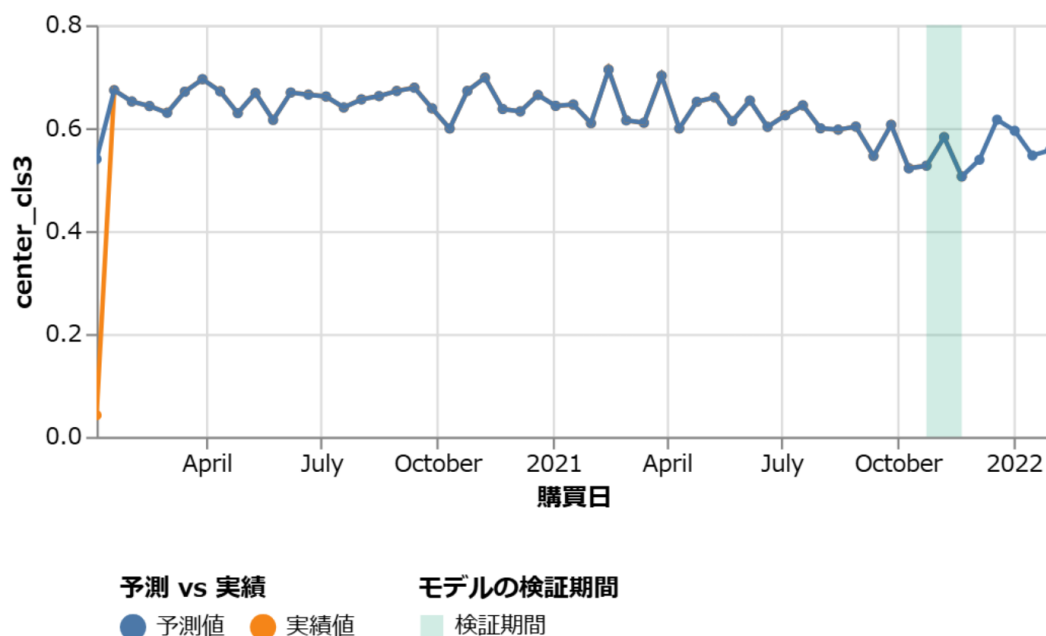
✓ 小売店のプロモーション施策が、**女性顧客の購買行動を促進できている**可能性があります。

アウトプット

Auto Modeling 将来予測

クラスタ中心線に対して時系列予測を行い、各グループの購買行動が今後どのように推移するかを予測します。ここでは、“cluster 3”（P.13で購買傾向が横ばいのクラスタ）のクラスタ中心線に対する時系列予測の結果を示します。

• クラスタ中心線の将来予測（prediction_plot）



プロットからの観察

- cluster 3のグループ全体の購買は、今後も横ばいの傾向を示すことが予測されます。

✓ 施策を実施しない場合でも、全体の購買傾向がこれまでと大きく変わらないと予測されるため、このグループへの**マーケティング施策の優先順位を下げる**といった判断が可能かもしれません。

アイコンの設定

プロジェクト 解説

時系列リサンプリング

再標本区間ごとの計算方法を指定します。

集計方法

再標本区間ごとの集計方法を指定します。

リサンプリング間隔

「年」・「月」・「週」・「日」から再標本間隔を選択します。例えば、「週」を選択すると週ごとのデータになるようにデータをまとめあげます。

ウィンドウサイズ

再標本化の間隔を正の整数で指定します。
例えば、再標本化の間隔を「週」、ウィンドウサイズを「2」と指定した場合、結果は2週間ごとに集約されます。

欠損値処理

等間隔になるよう再標本する際、その間隔内に値をもたない行を、欠損とみなすか指定します。欠損を補完する場合、固定値で補間したり、前後の値で補間できます。

時系列リサンプリング

変数選択	列名	列型	タイムスタン...	キー列	計算対象列
	顧客ID	整数	☐	☒	☐
	購買日	日付	☒	☐	☐
	購買個数.SUM	実数	☐	☐	☒
	購買金額.SUM	実数	☐	☐	☒

計算方法

集計方法*

合計

リサンプリング間隔*

週

ウィンドウサイズ

2

欠損値処理*

固定値

欠損値処理-固定値の場合の値

0

出力形式

出力形式*

縦持ち(ロング型)

☐ 期間指定を行う

期間指定-開始時刻

2021/01/02 00:00:01

期間指定-終了時刻

2022/12/31 00:00:00

プロジェクト 解説

時系列クラスタリング

時系列クラスタリングの計算方法を指定します。

クラスタ数

時系列を何個のクラスタに分けるかを指定します。

クラスタリング手法

時系列クラスタリングのアルゴリズムとして、「k-means」または「k-medoids」のいずれかを選択できます。

距離指標

時系列間の類似度（距離）をどのような指標で測るかを指定します。距離指標として、「euclidean」（ユークリッド距離）または「dtw」（Dynamic Time Warping 指標）のいずれかを選択できます。

正規化手法

時系列データの正規化の方法を指定します。正規化を行わない場合は「なし」、正規化を行う場合は「min-max」または「mean-std」のいずれかを選択します。

時系列クラスタリング

変数選択

列名	列型	タイムスタン...	キー列	計算対象列
顧客ID	整数	☐	☒	☐
購買日	日時	☒	☐	☐
購買個数 SUM	実数	☐	☐	☒
購買金額 SUM	実数	☐	☐	☐

計算方法 (基本設定)

クラスタ数
4

クラスタリング手法 *
k-means

距離指標 *
euclidean

正規化手法 *
min-max

計算方法 (詳細設定)

クラスター中心初期化方法 *
random

リストート数
10

反復回数
200

収束誤差
0.000001

本文書・プロジェクトファイルのご利用にあたって

本文書ならびにプロジェクトファイルは、株式会社NTT データ数理システム（以下「弊社」）が開発・販売する分析プラットフォーム **Alkano** についての情報提供として弊社が作成を行ったものです。弊社による事前の許可なしに、本文書の再配布や引用の範囲を超える複製といった行為、およびリバースエンジニアリングを禁じます。

本文書ならびにプロジェクトファイルのご利用に際して、ご利用者様および第三者に損害が発生したとしても、弊社は責任を負わないものとします。

プロジェクトファイルは、その中に同梱されているデータを利用し、本文書内で解説している設定可能なパラメータで動作させた場合についてのみ、弊社にて動作の検証を行っております。これを超えるような状況における動作は保証いたしません。

本プロジェクトファイル は、**MSIP1.10.0** および **Alkano1.4.0** にて動作確認を行っております。



データ活用の確かなパートナー

お問い合わせ・詳細:
株式会社NTTデータ数理システム
<https://www.msi.co.jp/solution/alkano/>

株式会社 NTTデータ数理システム

NTT DATA NTT DATA Mathematical Systems Inc.